

# Using Length of Stay to Control for Unobserved Heterogeneity When Estimating Treatment Effect on Hospital Costs with Observational Data: Issues of Reliability, Robustness, and Usefulness

*Peter May, Melissa M. Garrido, J. Brian Cassel, R. Sean Morrison, and Charles Normand*

---

**Objective.** To evaluate the sensitivity of treatment effect estimates when length of stay (LOS) is used to control for unobserved heterogeneity when estimating treatment effect on cost of hospital admission with observational data.

**Data Sources/Study Setting.** We used data from a prospective cohort study on the impact of palliative care consultation teams (PCCTs) on direct cost of hospital care. Adult patients with an advanced cancer diagnosis admitted to five large medical and cancer centers in the United States between 2007 and 2011 were eligible for this study.

**Study Design.** Costs were modeled using generalized linear models with a gamma distribution and a log link. We compared variability in estimates of PCCT impact on hospitalization costs when LOS was used as a covariate, as a sample parameter, and as an outcome denominator. We used propensity scores to account for patient characteristics associated with both PCCT use and total direct hospitalization costs.

**Data Collection/Extraction Methods.** We analyzed data from hospital cost databases, medical records, and questionnaires. Our propensity score weighted sample included 969 patients who were discharged alive.

**Principal Findings.** In analyses of hospitalization costs, treatment effect estimates are highly sensitive to methods that control for LOS, complicating interpretation. Both the magnitude and significance of results varied widely with the method of controlling for LOS. When we incorporated intervention timing into our analyses, results were robust to LOS-controls.

**Conclusions.** Treatment effect estimates using LOS-controls are not only suboptimal in terms of reliability (given concerns over endogeneity and bias) and usefulness (given the need to validate the cost-effectiveness of an intervention using overall resource use for a sample defined at baseline) but also in terms of robustness (results depend on the approach taken, and there is little evidence to guide this choice). To derive results that minimize endogeneity concerns and maximize external validity, investigators should match and analyze treatment and comparison arms on baseline factors only. Incorporo-

rating intervention timing may deliver results that are more reliable, more robust, and more useful than those derived using LOS-controls.

**Key Words.** Costs and cost analysis, palliative care, length of stay, endogeneity, validity

---

A randomized controlled trial is the strongest way to maximize internal validity when evaluating causal effects, but many interventions cannot be randomized, leaving health service researchers to analyze observational data (Black 1996; McKee et al. 1999).

The most significant weakness in observational designs is that, as assignment to treatment and comparison groups is not under analyst control, differences in outcome between subjects may result from the intervention, from observed confounders and from unobserved confounders (Stuart 2010; Austin 2011). Observed confounding can be controlled for by the use of statistical matching techniques, such as propensity scoring (Rosenbaum and Rubin 1983; Rubin 2007). An instrumental variable approach may be the best way to address unobserved confounding in a cross-sectional study (Angrist, Imbens, and Rubin 1996), but a strong, valid instrument is often not easy to identify, even in prospective studies (Murray 2006).

Unobserved heterogeneity is consequently a common concern for health services studies aiming to infer causal effects using observational data (Heller, Rosenbaum, and Small 2009). This concern is exacerbated in evaluation of treatment effect on utilization outcomes—for example, costs, number of readmissions—because these distributions are typically right-skewed by a minority of complex patients, who can distort treatment effect estimates even in the absence of confounding (Jones 2010; Mihaylova et al. 2011). Highly complex patients therefore represent a dual challenge in inferring treatment effects on utilization using observational data: they complicate matching procedures, as clinical complexity is difficult to capture adequately through either

---

Address correspondence to Peter May, Ph.D., Centre for Health Policy & Management, Trinity College Dublin, 3-4 Foster Place, Dublin 2, Ireland; e-mail: mayp2@tcd.ie. Peter May, Ph.D., is also with the Brookdale Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY. Melissa M. Garrido, Ph.D., and R. Sean Morrison, M.D., are with the James J. Peters VA Medical Center, Bronx, NY; Brookdale Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY. J. Brian Cassel, Ph.D., is with the Massey Cancer Center at Virginia Commonwealth University, Richmond, VA. Charles Normand, D.Phil., is with the Centre for Health Policy & Management, Trinity College Dublin, Dublin, Ireland.

administrative or interview data; and they skew the distributions of outcomes of interest, potentially obscuring important associations in the data.

Multiple studies evaluating the causal effect of a treatment on utilization during a hospital admission have addressed the problem of complex outliers by controlling for length of stay (LOS) as a proxy for clinical complexity. Specifically, one or more of three approaches have been variously employed:

1. *LOS as a covariate*: LOS, or a nonlinear transformation of LOS, included as a predictor in regression to estimate treatment effect on cost;
2. *LOS as a sample parameter*: Short- and/or long-stay outliers removed from the sample *ex ante*;
3. *LOS as an outcome denominator*: Average daily costs, that is, the ratio of total costs to LOS, employed as the primary outcome of interest.

Each of the three strategies is intended to address challenges in analysis of utilization data collected with an observational design: by controlling for LOS, investigators mitigate problems of unobserved complexity to the extent that LOS is a useful proxy for complexity and skewed utilization data to the extent that removing long-stay outliers or examining mean daily costs normalizes the distribution of the outcome of interest. But each strategy potentially undermines the internal and/or external validity of treatment effect estimates by increasing endogeneity and bias concerns, and by limiting the practical value of derived results. See Table 1 for an overview of these methods, their respective justifications and potential problems, and examples from studies of inpatient hospital utilization for maternity care, pediatric services, and palliative care. Example studies were all performed on the acute hospital setting, but the methodological principles are likely generalizable to other institutional settings for health care delivery. Evidence-based comparison of these strategies within a single study is limited.

### *Objectives*

The purpose of this paper was to compare and evaluate the sensitivity of estimates of a treatment's impact on hospital costs to methods of controlling for LOS, and to clarify the strengths and weaknesses of these methods.

The data are drawn from a prospective, observational study on the impact of palliative care consultation team (PCCT) interventions for hospitalized patients with advanced cancer. In this study, we explored methods that use LOS to control for unobserved heterogeneity. Our results and discussion

Table 1: Summary of Strategies That Control for LOS in Analysis of Hospital Utilization Using Observational Data

<i>Use of LOS</i>	<i>Definition—Potential Justifications</i>	<i>Potential Problems</i>	<i>Examples from Research</i>
I. Covariate	<p>LOS employed as an independent variable/predictor in regression analysis</p> <p>Covariate intended to control for:</p> <ul style="list-style-type: none"> <li>Unobserved heterogeneity, including hard-to-capture clinical complexity. LOS may be a useful proxy for complexity;</li> <li>Uneven accumulation of costs, higher costs being accrued early in hospitalization</li> </ul>	<p>Use of LOS as a covariate risks introducing endogeneity into analysis, as LOS is not an independent predictor of resource consumption or cost. Rather, it is associated with both treatment (long hospital stay suggests clinical complexity) and outcome (LOS and other utilization data are typically closely correlated), thus undermining estimation of the causal relationship of interest (Amorfu 2010; Garrido et al. 2012).</p>	<p>Kuo et al. (2012); Thompson et al. (2003); Whitford et al. (2014)</p>
II. Sample parameter	<p>Short- and/or long-stay outliers trimmed from sample</p> <p>Sample parameter(s) intended to control for:</p> <ul style="list-style-type: none"> <li>Unobserved heterogeneity, including clinical complexity. LOS outliers may be unrepresentative of the study sample in ways that are hard to capture—removing LOS outliers attempts to make the sample more homogenous;</li> <li>Outliers skewing distribution of utilization data such as LOS and cost, distorting and disguising treatment effects.</li> </ul>	<p>Defining a sample <i>ex ante</i> by a factor that is associated with both treatment and outcome risks biasing results and obscuring true treatment effect (Imbens 2004; Garrido 2014). Where propensity score matching is used*, <i>ex ante</i> trimming is antithetical to the research framework, which aims to estimate a counterfactual using baseline data (Rubin 2007). LOS is not known at admission, so evidence of treatment efficacy for a patient group defined by LOS is of limited practical use.</p>	<p>McCarthy et al. (2015); Morrison et al. (2008); Starks et al. (2013)</p>

Continued

Table 1: Continued

Use of LOS	Definition—Potential Justifications	Potential Problems	Examples from Research
III. Outcome denominator	<p>Average daily cost (the ratio of total cost to LOS) employed as primary outcome of interest</p> <p>Outcome denominator intended to:</p> <ul style="list-style-type: none"><li>• Indirectly limit impact of unobserved heterogeneity for which LOS is a potential proxy by accounting for long LOS;</li><li>• Reduce skew and leptokurtosis common to health care utilization data distributions;</li><li>• Address specific stakeholders, for example, a hospital reimbursed a fixed daily rate, who may prioritize average daily cost over total cost.</li></ul>	<p>Estimated effect on average daily costs is of limited practical value because this is a ratio and not overall resource use: a treatment that reduces daily cost by 10% but increases LOS by 50% (thus increasing total cost) is not necessarily delivering desirable clinical or financial outcomes (Weinstein et al. 1996). <i>Per diem</i> ratios change systematically with LOS and must be interpreted carefully (Ishak et al. 2012). If LOS differs between treatment and comparison groups, then daily cost (total cost/LOS) is a fundamentally different outcome to total cost.</p>	<p>Ciemiens et al. (2007); Penrod et al. (2010); Queen et al. (2014)</p>

*Note.* \*All examples cited in this table also used propensity scores to account for observed confounders, with the exceptions of Thompson and Ciemiens, who respectively used logistic regression and DRG matching to account for observed differences.

may be generalized to other observational studies of the impact of a binary intervention on utilization during an acute care admission.

In our primary analysis, we estimate the effect of a binary treatment variable on cost of hospital stay, controlling for LOS per Table 1: as a covariate, as a sample parameter, and as an outcome denominator. We compare the results and highlight considerations for analyses in a similar context. The derived results are highly sensitive to if and how LOS is controlled for. Treatment effect estimates using LOS-controls are not only suboptimal in terms of reliability, given concerns over endogeneity and bias; and usefulness, given the need to validate the cost-effectiveness of an intervention using overall resource use for sample defined at baseline; but also in terms of robustness, as results depend on the approach taken, and there is little evidence to guide this choice.

In the interpretation of our primary analysis, we discuss potential responses to these challenges. First, the importance of reporting sensitivity analyses is emphasized. Second, we show that an alternative method with our data—incorporating intervention timing during the hospital admission—improved markedly the rigor and consistency of our results compared to those derived using LOS-controls.

## METHODS

### *Sample*

We have reported elsewhere full details of our study consistent with the STROBE guidelines (May et al. 2015a).

Briefly, using a prospective, observational design, we collected descriptive, clinical, and utilization data on patients with an advanced cancer diagnosis admitted to one of five U.S. hospitals between 2007 and 2011 (the Palliative Care for Cancer [PC4C] study; National Institutes of Health, 2006). Patients were eligible if they had an advanced cancer diagnosis, were at least 18 years of age, spoke fluent English, and gave informed consent. Patients were excluded if they had a diagnosis of dementia, were nonverbal, or had previously received a PCCT intervention. Additional criteria included approval of the attending physician and a hospital stay of at least 48 hours. Clinical data came from medical record review and patient interviews. Cost data were extracted from hospital databases, adjusted for regional variation, and standardized to U.S. dollars (USD) in 2011, the final year of data collection. Of the five hospitals, one did not collect cost data and is excluded from this paper.

We used kernel weights derived from propensity scores to account for observed patient sociodemographic and clinical characteristics associated with both PCCT use and health care costs (Garrido et al. 2014). With kernel weighting, each subject in the treatment group is given a weighting of 1 and matched to a subject in the comparison group using a weighted average of the propensity score covariates. Only subjects outside the range of common support are dropped. Treatment and comparison groups were balanced on baseline covariates covering demographic, socioeconomic, clinical, and health system factors: age, gender, race, advance directive status, education level, insurance program, comorbidities, diagnosis, activity level, symptom scores, and access to formal homecare prior to admission. Balance was evaluated with standardized differences before and after weighting.

This analysis focuses on the 969 patients in our propensity score-weighted sample who were admitted to one of the four hospitals that provided cost data and discharged alive (May et al. 2015a). Dividing patients by discharge status is common in end-of-life care studies due to a possible heterogeneity problem: there are distinct implications of the point at which costs are no longer accrued as well as potential unobserved clinical complexity, and differences in treatment pathways and patient preferences (Cassel et al. 2010). For our study there were 54 patients who died, insufficient for a stand-alone weighted analysis. When patients who died are pooled with those discharged alive in our data, there are no substantive differences to any of our results. Where the sample was trimmed according to LOS, new sample-specific propensity scores were created accordingly: all observations outside the sample's LOS range were discarded and the weighting process repeated for the remaining treatment and comparison group patients (Green and Stuart 2014).

### *Variables*

The primary outcome of interest was total direct costs incurred by the hospital for the hospitalization (Taheri et al. 2000).

The primary independent variable was a binary treatment variable: Did patients receive a consultation from the PCCT during their hospital admission? Patients who were seen by a PCCT were given a score of 1 (i.e., the treatment [PC] group); those who were not were given a score of 0 (i.e., the comparison [UC] group). As with all medical consultations (e.g., oncology, cardiology, infectious disease), the PCCT intervention was initiated at the request of the primary treating physician when it was judged that patients with advanced cancer would benefit from specialist expertise. Reasons for consul-

tation included treatment of pain and other symptoms, clarifying treatment options, establishing goals of care and advance plans, and assistance with transition planning. The PCCT comprised a specialist-led interdisciplinary team of a physician, a nurse, and a social worker with chaplaincy and psychiatry support (American Academy of Hospice and Palliative Medicine, Center to Advance Palliative Care, Hospice and Palliative Nurses Association, Last Acts Partnership & National Hospice and Palliative Care Organization 2004).

Additional predictors were all variables included in the propensity score ( $n = 33$ ). All regressions were performed against the treatment variable and propensity score variables (random effects) plus fixed effects for each hospital site.

### *Statistical Methods*

Prior to our primary analysis, we compared the performance of different linear and nonlinear modeling approaches with Pearson, Hosmer–Lemeshow and link tests, and goodness of fit measures  $R^2$ , in-sample root mean squared error and in-sample mean absolute prediction error, consistent with Jones et al. (2013). A generalized linear model with a gamma distribution and log link was selected as the strongest option with these data. Details are available from the authors.

Using GLMs (gamma, log), we calculated the mean incremental effect of PCCT on total direct hospital costs. We performed sensitivity analyses using OLS models, and GLM (gamma, log) without propensity score weights, to establish that reported results were robust to model selection and weighting strategy. Incremental effects were calculated as the mean of finite differences across the sample and using bootstrapped standard errors (1,000 replications) (Abadie and Imbens 2008). All analyses were performed using *Stata* (version 12) (StataCorp, 2011).

### *LOS-Controls*

We estimated mean incremental treatment effect on total direct hospital costs using the three LOS-control strategies detailed in Table 1: LOS as a covariate, as a sample parameter, and as an outcome denominator.

*LOS as a Covariate.* LOS is linearly correlated with total cost of hospital stay, so the inclusion of an untransformed LOS covariate in the regression com-



pounds endogeneity concerns discussed in Table 1. To break the linear relationship between covariate and outcome, where health economists have incorporated LOS as a predictor in modeling costs, they have generally done so following nonlinear transformation: LOS log-transformed, squared, cubed, etc. (Carey and Burgess 1999; Penrod et al. 2010).

We therefore estimated five GLMs [Models (i), (ii), (iii), (iv), (v)] with a gamma distribution and a log link, differentiated by the use of common nonlinear transformations of LOS as a covariate. Our base model [Model (i)] included main effects for the treatment variable and the 33 observed covariates in the propensity score, fixed effects for site, and an error term. Models (ii), (iii), and (iv) included LOS to the second, third, and fourth power, respectively. Model (v) included a log-transformed LOS term.

*LOS as a Sample Parameter.* There is no established guideline for removing LOS outliers to improve the accuracy of analyses (Marazzi et al. 1998). For the summary statistics of LOS for our sample, see Table 2. We trimmed our sample of short-stay patients at the 5th percentile, creating a new subsample where  $4 \leq \text{LOS}$  ( $N = 942$ ); of long-stay patients at the 95th percentile ( $\text{LOS} \leq 20$ ;  $N = 924$ ), and of both simultaneously at the 2.5th and 97.5th percentiles ( $4 \leq \text{LOS} \leq 25$ ;  $N = 917$ ). These limits were not set according to any clinical parameters but to maximize sample size in investigating the impact of defining the sample by LOS.

For each subsample, we reran Models (i) to (v) and estimated mean incremental treatment effects.

*LOS as an Outcome Denominator.* Regressions were rerun with average daily direct cost, that is, the ratio of total direct hospital costs to LOS, as the outcome of interest for the whole sample and the three trimmed subsamples. These were calculated using only Model (i) (no LOS as a covariate) as LOS is already implicitly incorporated in the regression via the outcome of interest.

## RESULTS

Treatment effect estimates using each of the three LOS-control strategies are given in Table 3. Results differed substantively depending on whether and how LOS was controlled for in analyses.

Table 2: Selected Statistics of Principal Sample ( $N = 969$ )

	Before Matching			After Matching		
	UC ( $n = 713$ )	PC ( $n = 256$ )	All Patients ( $N = 969$ )	UC	PC	Standardized Difference
Age						
<55	215 (30.2%)	91 (35.6%)	306 (31.6%)	35.4%	35.5%	-0.3%
55-75	416 (58.4%)	137 (53.5%)	553 (57.1%)	51.6%	53.5%	-3.8%
75<	82 (11.5%)	28 (10.9%)	110 (11.4%)	13.0%	10.9%	6.3%
Gender						
Female	400 (56.1%)	137 (53.5%)	537 (55.4%)	55.0%	53.5%	3.0%
Race						
White	490 (68.7%)	158 (61.7%)	648 (66.9%)	61.2%	61.7%	-1.0%
Black	170 (23.8%)	85 (33.2%)	255 (26.3%)	33.5%	33.2%	0.6%
Other	53 (7.4%)	13 (5.1%)	66 (6.8%)	5.3%	5.1%	1.0%
Insurance						
Medicare	129 (18.1%)	50 (19.5%)	179 (18.5%)	19.5%	19.5%	0.0%
Medicaid	102 (14.3%)	65 (25.4%)	167 (17.2%)	25.8%	25.4%	0.9%
Other	482 (67.6%)	141 (55.1%)	623 (64.3%)	54.7%	55.1%	-0.7%
Advance directive						
Yes	429 (60.2%)	130 (50.8%)	559 (57.7%)	50.1%	50.8%	-1.3%
Primary diagnosis						
Lymphoma	61 (8.6%)	15 (5.9%)	76 (7.8%)	6.1%	5.9%	1.0%
Comorbidities						
Elixhauser total	3.22	3.96	3.41	3.94	3.96	-1.1%
Activity level						
ADL-6	10.76	9.8	10.50	9.82	9.78	1.5%
Symptoms						
ESAS physical	1.45	1.97	1.59	1.98	1.97	1.2%

Continued

Table 2: Continued

	Before Matching			After Matching		
	UC (n = 713)	PC (n = 256)	All Patients (N = 969)	UC	PC	Standardized Difference
ESAS psych.	1.45	1.61	1.49	1.57	1.61	−2.8%
Total cost (\$1,000)						
Mean	9.55	11.15	9.97			
25th/50th/75th%	4.9/7.4/11.8	4.8/7.4/12.3	4.9/7.4/11.8			
LOS (days)						
Mean	8.0	9.0	8.2			
25th/50th/75th%	5/6/9	5/7/10	5/7/9			

Note. Medicare: Patients with Medicare and no other insurance; Medicaid: Patients with Medicaid.

Table 3: Estimated Treatment Effect on Direct Hospital Costs (in US\$ with 95% CI) Using Different LOS-Control Strategies

	All (N = 969; UCn = 713, PCn = 256)	4≤LOS (N = 942; UCn = 713, PCn = 229)	4≤LOS≤25 (N = 917; UCn = 717, PCn = 200)	LOS≤20 (N = 924; UCn = 686, PCn = 238)
Model (i) [No LOS covariate]	+153 (−1,266 to +1,572)	+773 (−1,307 to +2,853)	−438 (−1,483 to +607)	−1,141 (−1,983 to −299)*
Model (ii) [LOS^2 covariate]	−1,597 (−2,366 to −829)*	−1,423 (−2,249 to −596)*	−1,432 (−2,120 to −744)*	−1,595 (−2,227 to −964)*
Model (iii) [LOS^3 covariate]	−1,617 (−2,438 to −796)*	−1,358 (−2,773 to +56)	−1,341 (−2,067 to −614)*	−1,609 (−3,134 to −84)*
Model (iv) [LOS^4 covariate]	−1,549 (−2,408 to −691)*	−1,253 (−2,159 to −347)*	−1,235 (−1,989 to −480)*	−1,565 (−2,282 to −849)*
Model (v) [ln(LOS) covariate]	−980 (−1,701 to −261)*	−1,193 (−2,357 to −30)*	−1,465 (−2,076 to −855)*	−1,148 (−2,228 to −68)*
Daily cost as outcome of interest	−110 (−195 to −24)*	−174 (−249 to −98)*	−135 (−227 to −82)*	−141 (−227 to −54)*
[No LOS covariate]				

\*  $p < .05$ ; \*  $p < .01$ .

*LOS as a Covariate*

Within each sample, results differed substantively, depending on if and how LOS is included as a predictor in regression.

For the principal sample ( $N = 969$ ), Model (i) (no LOS covariate) suggested a negligible association between treatment and costs: +\$153 [95 percent CI:  $-1,266$  to  $+1,572$ ], equivalent to a 1.6 percent increase in total direct hospital costs for patients who received PC ( $p = .83$ ). When LOS was included as a covariate (regardless of how it was transformed), estimated treatment effects were significant and consistently suggested cost-savings from the intervention.

Similarly, for the samples truncated of short-stayers, and short- and long-stayers simultaneously, nonsignificant estimates using Model (i) contrasted with statistically significant cost-savings with Models (ii) to (v) in the \$980–\$1,617 range, equivalent to a 10–17 percent cost-saving from the intervention.

Only for the sample with long-stayers removed was there broad consistency between all Models: a statistically significant cost-saving association is reported for each. But even in this case there is variation in the magnitude of the projected saving, from \$1,141 to \$1,609, meaning that the largest estimate is 41 percent larger than the smallest.

*LOS as a Sample Parameter*

A comparison of results for Model (i—no LOS as a covariate) shows that estimates differ substantively depending on whether and how the sample is trimmed by LOS.

For the principal sample ( $N = 969$ ), the estimated treatment effect was negligible, but when the sample was truncated by short-stay outliers only, or both short-stay and long-stay outliers, the estimated treatment effect was not statistically significant ([95 percent CI:  $-1,307$  to  $+2,853$ ;  $p = .47$ ] and [95 percent CI:  $-1,483$  to  $+607$ ;  $p = .41$ ], respectively).

However, trimming long-stay outliers resulted in a statistically significant mean cost-saving effect of  $-\$1,141$  (95 percent CI:  $-1,983$  to  $-299$ ;  $p = .008$ ), equivalent to a 12 percent reduction in total direct hospital costs per case.

*LOS as an Outcome Denominator*

The estimated treatment effects on average daily direct cost (total direct hospital costs/LOS) suggest a statistically significant cost-saving effect irrespective

of sample or subsample in the range \$110–\$174. Again, this represents a large variation (58 percent) between the smallest and largest estimates. The estimated effects on mean daily costs are equivalent to 9–13 percent reductions, slightly lower than the equivalent estimated savings where LOS was employed as a covariate.

## DISCUSSION

The approach chosen in incorporating LOS in analysis has a direct bearing on results (Table 3). The estimated effect of treatment on direct hospital costs without controlling for LOS is negligible. The estimated effect once LOS is employed as a covariate, sample parameter, or outcome denominator is generally statistically significant and cost-saving, with substantive variance in the magnitude of estimated effect.

Previous analyses of the impact of palliative care and other hospital-based interventions on utilization using observational data have variously used different approaches to LOS to control for unobserved heterogeneity, often in tandem with propensity scoring to control for observed confounders. But our results raise a number of considerations in evaluating treatment effect on hospital costs and, implicitly, other utilization outcomes using these methods. Specifically, there are issues of reliability, robustness, and usefulness.

First, strategies to control for LOS are suboptimal in research design. Use of LOS as a covariate introduces unquantifiable endogeneity into the analysis as the duration of a patient's hospital stay is strongly correlated with the cost of that stay. Use of LOS as a sample parameter loses information, weakens power and undermines the fundamentals of the propensity score methods, which allow calculation of a counterfactual on the basis of observed covariates at baseline. Treatment effect estimates on average daily cost are difficult to interpret if LOS differs between treatment and comparison group, as is the case with our data. Estimates derived using these methods may not be reliable.

Second, there is little guidance on how or when to incorporate LOS as a covariate to control for unobserved heterogeneity, or at what levels to trim the sample of short- and long-stay outliers. Yet if and how such methods are employed has a direct bearing on results. As this is an observational dataset of 969 patients, it is not possible to know the “true” treatment effect but, given the lack of observable association between treatment and cost when LOS is

not controlled for, accepting results using LOS-controls may increase the risk of a type I error. There is no treatment effect reported in Table 3 that could be confidently reported as robust to sensitivity analysis.

Third, results calculated using LOS-controls have limited practical use. Evidence that a treatment is efficacious for a sample defined by LOS cannot directly inform policy or clinical practice because LOS is not known at admission. Similarly, estimated treatment effect on average daily cost may be misleading because average daily cost does not reflect overall resource consumption or cost to the hospital; it is merely a ratio. There is a well-known maxim in health economics that investigators should not estimate treatment effect on log-transformed costs because this is fundamentally not useful: “Congress does not appropriate log dollars. First Bank will not cash a check for log dollars” (Manning 1998). This principle could be extended to employing LOS-controls in estimating impact of a binary treatment variable on costs of a hospital admission: interventions are not validated as (cost-)effective for populations defined *ex ante* by LOS, or on the basis that they reduce the ratio of total cost to LOS during hospital stay. Gold-standard guidelines in health economics study design recommend that investigators estimate effect on overall utilization in the outcome of interest for a sample defined at baseline (Gold et al. 1996; Drummond et al. 2005).

### *Interpretation*

Our results therefore present challenges in interpretation. In particular, the association between the intervention and hospital costs is not clear. The significance and magnitude of the estimated effect are dictated by if and how LOS is controlled for analysis, and little guidance exists to inform a specific approach.

This suggests that multiple sensitivity analyses with and without LOS-controls, which are not typically reported in studies employing these methods to date, ought to be reported as standard when employing methods of this type. The true treatment effect can never be known in analysis of observational data, but where results are robust to sensitivity analyses investigators can be increasingly confident that the associations they report approximate the true effect. Conversely, where results are highly sensitive, their value is necessarily limited.

However, consistency between results generated using LOS-controls is reassuring only to the extent that LOS is a useful proxy for some important unobserved factor(s). Concerns about endogeneity, bias, and usefulness remain for these methods, no matter how consistent the results may be in sen-

sitivity analyses. Optimally, treatment effect estimates should not only be robust to sensitivity analyses but also based on samples defined and data measured at baseline.

*An Alternative to LOS-Controls: Intervention Timing*

Our results (Table 3) suggest that LOS is indeed a useful proxy for some unobserved factor. The fact that almost any use of LOS-control leads to a statistically significant association between treatment and cost, yet the absence of any such control finds no association at all, indicates that there is a treatment effect for a latent class in the sample and LOS is an indicator of this latent class. Accepting that there is a negligible association would therefore miss important relationships evident in the data. In particular, it appears that a small number of long-stay patients are masking a treatment effect for the majority: this treatment effect becomes visible if LOS is controlled for as a covariate, if long-stay outliers are removed, or if average daily costs are used as an outcome of interest.

To estimate treatment effect on cost with our data, this prompted the question: How could the latent heterogeneity be meaningfully identified using input data rather than output? In particular, what defined long-stay patients other than LOS and could this be meaningfully identified using input/baseline data?

We subsequently discovered that time-to-consult is an influential factor in determining treatment effect on cost: early palliative care is associated with a significant cost-saving effect, and this effect is larger when treatment is received earlier (May et al. 2015a).

Early palliative care is also shown to reduce LOS. This association requires further investigation, but we hypothesize that this reflects goals of care discussions and transition planning through the intervention, leading to fewer high-intensity treatments and earlier discharge.

Incorporating intervention timing into analyses is not a direct substitute for LOS-control strategies because it does not solve the problem of unobserved variation in clinical complexity. However, it does offer important advantages over LOS-controls in these types of analyses.

First, it delivers a more accurate definition of the intervention. Long-stay patients systematically have later consults and costs are a cumulative outcome of interest. All patients accrue costs from the point of admission that are included in the cost of hospital stay; for late-consult patients, a higher proportion of these costs are by definition not amenable to treatment.



The relevant data are presented in Table 4. Intervention timing is systematically associated with the proportion of the outcome of interest that treatment can affect: patients who receive an intervention within 2 days of admission have accrued on average 28 percent of costs prior to treatment; for later interventions the equivalent figure is 61 percent. Moreover, late-consult patients are systematically higher cost patients, meaning that when intervention timing is not considered, these patients have a *larger* impact on mean cost estimates, even though their treatment is *less* likely to have a chance to affect the outcome of interest. Patients who receive a consult late in their hospital stay are therefore less representative of the intervention studied, yet, as high-utilization outliers, may be more influential than other patients on the treatment effect estimated.

Where the timing of the intervention may vary during the hospitalization, the assumption of homogenous treatment effect on hospital costs irrespective of timing in such analyses will therefore typically be false, increasing the risk of a type II error. It is also plausible that more clinically complex patients are more likely to receive a consultation earlier, so some degree of unobserved clinical complexity may be picked up by timing-sensitive definitions of treatment.

Second, after improving the specification of the intervention by incorporating timing, the derived treatment effect estimates are much more consistent for LOS-controls. Treatment effect estimates for an early intervention, for different samples defined by LOS, are given in Table 5. In contrast to estimates where intervention timing is not included, or where an LOS covariate is added, the estimated treatment effect is consistent and variability in the magnitudes of cost-saving effect is markedly reduced. The results are also more consistent for other approaches to long-stay, high-cost, and late-treatment outliers (details from authors).

The results of analyses incorporating intervention timing (Tables 4 and 5) also contribute substantively to our understanding of the inconsis-

Table 4: Utilization Summary, by Time-to-Intervention ( $n = 226$ )\*

<i>t</i> Days to First PCCT	PC (n)	Mean LOS (Days)	Mean Total Direct Costs (\$)	Mean Total Direct Costs Prior to First Consultation (\$)	Proportion of Costs Incurred Prior to First Consultation
$2 < t$	49	15	20,065	12,032	61%
$t \leq 2$	177	7	8,692	2,072	29%

Note. \*One site, with 30 PC patients, did not collect cost data in this way. PC group for this analysis is therefore 226, not 256, per the primary analysis.

Table 5: Estimated Treatment Effect on Total Direct Hospital Costs (in US\$ with 95% CI) for Different LOS-Defined Subsamples, Where Intervention Is Defined According to Timing and LOS Is Not Employed as a Covariate

	<i>All</i>	$4 \leq LOS$	$4 \leq LOS \leq 25$	$LOS \leq 20$
Intervention at any time (per Table 3)	+153 (−1,266 to +1,572)	+773 (−1,307 to +2,853)	−438 (−1,483 to +607)	−1,141 (−1,983 to −299)*
Intervention within 2 days	−2,280 (−3,438 to −1,122)*	−2,053 (−3,128 to −977)*	−1,767 (−2,840 to −693)*	−2,316 (−3,197 to −1,434)*

\* $p < .01$ .

tency in results using LOS-controls (Table 3). Specifically, a comparison suggests that in part LOS as a covariate was a proxy for intervention timing. Table 4 shows that later treatment is systematically associated with longer LOS. In the top row of Table 5, only where long-stay outliers are removed is there a significant treatment effect. In the bottom row of Table 5, once timing is taken into account, the treatment effect is much more consistent across all subsamples. And in Table 3, adding LOS as a covariate delivers a statistically significant effect for all subsamples in nearly all cases. Therefore, we can infer that the removal of long-stay patients is indirectly the removal of late-consult outliers, and controlling for LOS as a covariate indirectly controls for treatment timing. Because incorporating timing minimizes the endogeneity and bias concerns attendant to LOS as a covariate, as well as reducing the chance of a type II error by employing a binary any-time treatment variable for a cumulative outcome variable, we can be more confident in the reported association between an early intervention and hospital costs in Table 5, than in the corresponding estimate in Table 3.

### Limitations

There may be unobserved baseline heterogeneity between treated and comparison patients that is unconnected to LOS and treatment timing, and is therefore unaccounted for in our analyses. We have performed sensitivity analyses to confirm that results are robust to both regression model and weighting strategy. Results are also robust to model specifications with fewer covariates.

This is a single empirical study, and results need to be replicated in other empirical datasets. In addition, simulation studies would allow us to quantify the degree to which bias is exacerbated in each of the different methods, given a set of investigator-specified data-generating processes. However, the results

for this study provide an empirical illustration of the variability in treatment effect estimates that can arise from different methods of controlling for LOS in health services utilization studies. Investigators should be aware of the potential for variability in estimates and the need to run careful sensitivity analyses when examining the effect of treatments among hospitalized patients with varying lengths of stay.

The issues raised in this paper are primarily applicable for the estimation of treatment effect on costs during an acute hospital admission. Our results are robust in our data to inclusion or exclusion of patients who died during hospitalization, to LOS-controls, and to other measures of complexity such as number of comorbidities, but it is not clear that this will apply in all studies. While LOS will vary in its strength as a proxy for complexity, the issues of endogeneity, bias, and usefulness ought to be considerations for estimating treatment effect on utilization in all settings.

Incorporating treatment timing in the manner we suggest is a relevant strategy only in analysis of an intervention that is provided for the first time to different patients at different times during their hospitalization; it will not be relevant for an intervention that is provided either at admission or not at all. There are multiple options for incorporating timing; published examples include defining treatment as occurring within a specified time frame and either excluding patients who received the intervention after the specified time frame from analyses (May et al. 2015a) or including them in the comparison group (May et al. 2015b), and including interaction terms specifying timing in the regression model (McCarthy et al. 2015).

Our analyses may not be directly relevant to other uses of LOS in health services research, for example, as a proxy for, *inter alia*, quality control, appropriateness of care, resource consumption, and efficiency (Marazzi et al. 1998; Carey and Burgess 1999). In addition, it has been argued that truncating samples of high-utilization outliers is a valid approach, provided this decision is guided by clear scientific principles to improve model fit and the accuracy of treatment effect estimates (Marazzi et al. 1998; Mihaylova et al. 2011).

## CONCLUSION

Use of LOS-controls in estimating treatment effect on costs of hospitalization is at best suboptimal.

If and how LOS-controls are incorporated into cost analysis of observational data may have a substantial effect on results. These strategies risk under-

mining confidence in the treatment effect estimate by weakening key elements of the methodology and analysis. Employment of these strategies is sometimes arbitrary and little clear guidance exists to inform selection.

Nonetheless unobserved heterogeneity remains a serious concern for investigators analyzing treatment effect on cost and other utilization outcomes using observational data for hospital admissions. In our study a comparison of treatment effect estimates using LOS strategies suggested unobserved heterogeneity for which we would like to control.

We addressed this challenge by incorporating the timing of the intervention into our analyses. This approach does not solve directly problems of unobserved baseline heterogeneity, but it is scientifically preferable to LOS-controls. First, in avoiding the introduction of LOS, which is associated with both treatment and outcome, such methods generate treatment effect estimates that minimize the endogeneity problem. Second, in testing these results with multiple sensitivity analyses including LOS-controls, results are more robust as effects are not sensitive to factors for which LOS may be a proxy. Third, the results are more useful to hospital administrators, payers, and policy makers because patients are not defined *ex ante* by LOS, and the outcome of interest is the most important in cost analysis, namely overall resource use.

Additionally, they address a problem that has not been well identified in previous analyses of this type: that failure to incorporate intervention timing in analysis of treatment effect on cost of a hospital admission may increase the risk of a type II error.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* The authors thank Diane Meier, Thomas Smith, Robert Arnold, Phil Santa Emma, Mary Beth Happ, Tim Smith, and David Weissman for their contributions to the “PC4C” project. The PC4C study was funded by the National Cancer Institute (NCI) and the National Institute of Nursing Research in the United States (5R01CA116227-04). Dr. May is sponsored by a Health Economics Fellowship from the Health Research Board (Ireland) and NCI. Dr. Garrido is supported by a Veterans Affairs Health Services Research and Development career development award (CDA 11-201/CDP 12-255). This work was supported by the NIA, Claude D. Pepper Older Americans Independence Center at the Icahn School of Medicine at Mount Sinai [5P30AG028741], and the National

Palliative Care Research Center. Dr. Morrison was the recipient of a Midcareer Investigator Award in Patient-Oriented Research (5K24AG022345) during the course of this work. Sponsors had no role in design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript. All authors are independent of the study sponsors.

*Disclosures:* None.

*Disclaimers:* The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

## REFERENCES

- Abadie, A., and G. W. Imbens. 2008. "Notes and Comments on the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76: 1537–57.
- American Academy of Hospice and Palliative Medicine, Center to Advance Palliative Care, Hospice and Palliative Nurses Association, Last Acts Partnership & National Hospice and Palliative Care Organization 2004. "National Consensus Project for Quality Palliative Care: Clinical Practice Guidelines for Quality Palliative Care, Executive Summary." *Journal of Palliative Medicine* 7: 611–27.
- Amporfufu, E. 2010. "Estimating the Effect of Early Discharge Policy on Readmission Rate: An Instrumental Variable Approach." *Health* 2: 504–10.
- Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of American Statistical Association* 91: 444–55.
- Austin, P. C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46: 399–424.
- Black, N. 1996. "Why We Need Observational Studies to Evaluate the Effectiveness of Health Care." *British Medical Journal* 312: 1215–8.
- Carey, K., and J. F. Burgess Jr. 1999. "On Measuring the Hospital Cost/Quality Trade-Off." *Health Economics* 8: 509–20.
- Cassel, J. B., K. Kerr, S. Pantilat, and T. J. Smith. 2010. "Palliative Care Consultation and Hospital Length of Stay." *Journal of Palliative Medicine* 13: 761–7.
- Ciemiens, E. L., L. Blum, M. Nunley, A. Lasher, and J. M. Newman. 2007. "The Economic and Clinical Impact of an Inpatient Palliative Care Consultation Service: A Multifaceted Approach." *Journal of Palliative Medicine* 10: 1347–55.
- Drummond, M. F., M. J. Sculpher, G. W. Torrance, B. J. O'Brien, and G. L. Stoddart. 2005. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd Edition. Oxford, UK: Oxford University Press.
- Garrido, M. M. 2014. "Propensity Scores: A Practical Method for Assessing Treatment Effects in Pain and Symptom Management Research." *Journal of Pain and Symptom Management* 48: 711–8.

- Garrido, M. M., P. Deb, J. F. Burgess Jr., and J. D. Penrod. 2012. "Choosing Models for Health Care Cost Analyses: Issues of Nonlinearity and Endogeneity." *Health Services Research* 47: 2377–97.
- Garrido, M. M., A. S. Kelley, J. Paris, K. Roza, D. E. Meier, R. S. Morrison, and M. D. Aldridge. 2014. "Methods for Constructing and Assessing Propensity Scores." *Health Services Research* 49: 1701–20.
- Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- Green, K. M., and E. A. Stuart. 2014. "Examining Moderation Analyses in Propensity Score Methods: Application to Depression and Substance Use." *Journal of Consulting and Clinical Psychology* 82: 773–83.
- Heller, R., P. R. Rosenbaum, and D. S. Small. 2009. "Split Samples and Design Sensitivity in Observational Studies." *Journal of the American Statistical Association* 104: 1090–101.
- Imbens, G. W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* 86: 4–29.
- Ishak, K. J., M. Stolar, M. Y. Hu, P. Alvarez, Y. Wang, D. Getsios, and G. C. Williams. 2012. "Accounting for the Relationship between Per Diem Cost and LOS When Estimating Hospitalization Costs." *BMC Health Services Research* 12: 439.
- Jones, A. M. 2010. *Models for Health Care*. Working Paper 10/01. York, UK: Health Economics and Data Group, University of York.
- Jones, A. M., N. Rice, T. Bago D'Uva, and S. Balia. 2013. "Chapter 12: Modelling Health Care Costs." In *Applied Health Economics*, 2nd Edition, pp. 342–80. Oxford, UK: Routledge.
- Kuo, D. Z., L. L. Sisterhen, T. E. Sigrest, J. M. Biazio, M. E. Aitken, and C. E. Smith. 2012. "Family Experiences and Pediatric Health Services Use Associated with Family-Centered Rounds." *Pediatrics* 130: 299–305.
- Manning, W. G. 1998. "The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem." *Journal of Health Economics* 17: 283–95.
- Marazzi, A., F. Paccaud, C. Ruffieux, and C. Beguin. 1998. "Fitting the Distributions of Length of Stay by Parametric Models." *Medical Care* 36: 915–27.
- May, P., M. M. Garrido, J. B. Cassel, A. S. Kelley, D. E. Meier, C. Normand, T. J. Smith, L. Stefanis, and R. S. Morrison. 2015a. "Prospective Cohort Study of Hospital Palliative Care Teams for Inpatients with Advanced Cancer: Earlier Consultation Is Associated with Larger Cost-Saving Effect." *Journal of Clinical Oncology* 33: 2745–52.
- . 2015b. "Prospective Cohort Study of Hospital Palliative Care Teams for Inpatients with Advanced Cancer: Earlier Consultation Is Associated with Larger Cost-Saving Effect." *Journal of Clinical Oncology* [Online Data Supplement]. Epub 2015/06/08.
- McCarthy, I. M., C. Robinson, S. Huq, M. Philastre, and R. L. Fine. 2015. "Cost Savings from Palliative Care Teams and Guidance for a Financially Viable Palliative Care Program." *Health Services Research* 50: 217–36.
- McKee, M., A. Britton, N. Black, K. McPherson, C. Sanderson, and C. Bain. 1999. "Methods in Health Services Research. Interpreting the Evidence: Choosing

- between Randomised and Non-Randomised Studies." *British Medical Journal* 319: 312–5.
- Mihaylova, B., A. Briggs, A. O'Hagan, and S. G. Thompson. 2011. "Review of Statistical Methods for Analysing Healthcare Resources and Costs." *Health Economics* 20: 897–916.
- Morrison, R. S., J. D. Penrod, J. B. Cassel, M. Caust-Ellenbogen, A. Litke, L. Spragens, and D. E. Meier. 2008. "Cost Savings Associated with US Hospital Palliative Care Consultation Programs." *Archives of Internal Medicine* 168: 1783–90.
- Murray, M. P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspective* 20: 111–32.
- National Institutes of Health. 2006. *Project Information: Palliative Care for Hospitalized Cancer Patients [Online]*. Bethesda, MD: US Department of Health & Human Services [accessed on August 1, 2015]. Available at [http://projectreporter.nih.gov/project\\_info\\_description.cfm?projectnumber=5R01CA116227-04](http://projectreporter.nih.gov/project_info_description.cfm?projectnumber=5R01CA116227-04)
- Penrod, J. D., P. Deb, C. Dellenbaugh, J. F. Burgess Jr., C. W. Zhu, C. L. Christiansen, C. A. Luhrs, T. Cortez, E. Livote, V. Allen, and R. S. Morrison. 2010. "Hospital-Based Palliative Care Consultation: Effects on Hospital Cost." *Journal of Palliative Medicine* 13: 973–9.
- Queen, M. A., A. L. Myers, M. Hall, S. S. Shah, D. J. Williams, K. A. Auger, K. E. Jerardi, A. M. Statile, and J. S. Tieder. 2014. "Comparative Effectiveness of Empiric Antibiotics for Community-Acquired Pneumonia." *Pediatrics* 133: e23–9.
- Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55.
- Rubin, D. B. 2007. "The Design versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials." *Statistics in Medicine* 26: 20–36.
- Starks, H., S. Wang, S. Farber, D. A. Owens, and J. R. Curtis. 2013. "Cost Savings Vary By Length of Stay for Inpatients Receiving Palliative Care Consultation Services." *Journal of Palliative Medicine* 16: 1215–20.
- StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.
- Stuart, E. A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25: 1–21.
- Taheri, P. A., D. Butz, L. C. Griffes, D. R. Morlock, and L. J. Greenfield. 2000. "Physician Impact on the Total Cost of Care." *Annals of Surgery* 231: 432–5.
- Thompson, A. H., A. Alibhai, L. D. Saunders, D. C. Cumming, and N. Thanigasalam. 2003. "Post-Maternity Outcomes Following Health Care Reform in Alberta: 1992-1996." *Canadian Journal of Public Health* 94: 104–8.
- Weinstein, M. C., J. E. Siegel, M. R. Gold, M. S. Kamlet, and L. B. Russell. 1996. "Recommendations of the Panel on Cost-Effectiveness in Health and Medicine." *Journal of the American Medical Association* 276: 1253–8.
- Whitford, K., N. D. Shah, J. Moriarty, M. Branda, and B. Thorsteinsdottir. 2014. "Impact of a Palliative Care Consult Service." *American Journal of Hospice and Palliative Care* 31: 175–82.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.